ASA News
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®
732 North Washington Street, Alexandria, VA 22314 • (703) 684-1221 • Toll Free: (888) 231-3473 • www.amstat.org • www.twitter.com/AmstatNews

New Method Using Publicly Available Statistics Outperforms Other March Madness Upset Prediction Techniques

Novel method combines cutting-edge computational and statistical approaches

ALEXANDRIA, Va. (March 7, 2018) – University of Illinois researchers have developed a method using causal inference for predicting upsets in the NCAA Men's Basketball Tournament that outperforms many other techniques. In addition to improved accuracy, the method stands out because it relies on publicly available data, making it reproducible and more accessible for others to use.

The paper reporting the method is published in the American Statistical Association (ASA) *Journal of Quantitative Analysis in Sports* (JQAS) by Sheldon H. Jacobson (University of Illinois at Urbana-Champaign), Jason J. Sauppe (University of Wisconsin La Crosse) and Shouvik Dutta (former University of Illinois graduate student). In short, the technique identifies potential upsets using only a small number of publicly available statistics by identifying match-ups in the current year that exhibit characteristics similar to those exhibited by historical round-of-64 upsets.

Using decision trees, machine learning, and causal inference, Jacobson and his collaborators analyzed 115 publicly available statistics to detect the 15 most important for identifying upsets in the first-round matchups between the teams seeded 2 and 15, 3 and 14, and 4 and 13. Among the most influential of the 15 were the effective possession ratio—the number of possessions and offensive rebounds minus the number of turnovers all divided by the number of possessions—the number of games played in the regular season and a measure of scoring chances per game.

The differences in those 15 statistics between the two teams in each historical upset are then used to build a profile of past upsets. Finally, the upset profiles can be compared to round-of-64 games in the current year to find match-ups that are most like historical upsets.

Jacobson and co-authors applied their approach to the NCAA tournament in each of the 13 years from 2003 to 2015. Of the 26 selected games, 10 (38.4%) were actual upsets, which is more than twice as many as the expected number of correct selections when using a weighted random selection method.

Identifying causal factors in the NCAA tournament is challenging for many reasons, one being that randomized controlled trials—an established method ideally suited for identifying causality—is not an option. “By approaching the problem as a causal inference problem using observational data,” said Jacobson, “we were able to improve on forecasting upsets over pure random chance. “

Dubbed balance optimization subset selection (or BOSS), the framework can be applied to a broad array of data in the social sciences and medicine. The initial research for the BOSS idea was supported in part by the National Science Foundation. “The covariate balance approach taken by the authors is novel in the context of a sports application,” said Mark Glickman (Harvard University), former editor-in-chief of *JQAS* who handled this manuscript. “It is refreshing to see causal inference play a prominent role in assessing factors that impact game upsets.”

Jacobson's projected upsets for this year's tournament will be posted after Selection Sunday at http://bracketodds.cs.illinois.edu, a STEM learning laboratory focused on the statistics of March Madness.

"March Madness is a superb opportunity for all people, young and old, to enjoy a national sporting event while gaining an appreciation for how statistics and data science shed light on the tournament. Simply put, our research program on data analysis helps makes sense of the madness," said Jacobson.

Jacobson is a judge in the second annual Statsketball contest, hosted by *This*IsStatistics (http://thisisstatistics.org), the ASA's campaign to make students, teachers and parents aware of the many careers empowered by statistical thinking.

Copies of the paper are freely available to reporters by contacting the ASA.

###

About the American Statistical Association

The ASA is the world's largest community of statisticians and the oldest continuously operating professional science society in the United States. Its members serve in industry, government and academia in more than 90 countries, advancing research and promoting sound statistical practice to inform public policy and improve human welfare. For additional information, please visit the ASA website at www.amstat.org.

For more information:

Steve Pierson

ASA Director of Science Policy

(703) 302-1841

pierson@amstat.org

Sheldon H. Jacobson, PhD

Founder Professor of Computer Science

Department of Computer Science

University of Illinois at Urbana-Champaign

(217) 244-7275

shj@illinois.edu

twitter @shjanalytics